

PCTWELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales BüroINTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE
INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)(51) Internationale Patentklassifikation ⁶ :

G06F 17/30

A1

(11) Internationale Veröffentlichungsnummer: WO 97/36250

(43) Internationales
Veröffentlichungsdatum:

2. Oktober 1997 (02.10.97)

(21) Internationales Aktenzeichen: PCT/EP97/01442

(22) Internationales Anmeldedatum: 21. März 1997 (21.03.97)

(30) Prioritätsdaten:

196 11 672.4 25. März 1996 (25.03.96) DE
196 35 351.3 31. August 1996 (31.08.96) DE(71)(72) Anmelder und Erfinder: WADEWITZ, Rolf [DE/DE];
Ruhr-Möhne-Eck 9, D-59755 Amsberg (DE).

(81) Bestimmungsstaaten: US, europäisches Patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Veröffentlicht

*Mit internationalen Recherchenbericht.
Vor Ablauf der für Änderungen der Ansprüche zugelassenen
Frist. Veröffentlichung wird wiederholt falls Änderungen
eintreffen.*

(54) Title: DATA-CONVERSION METHOD

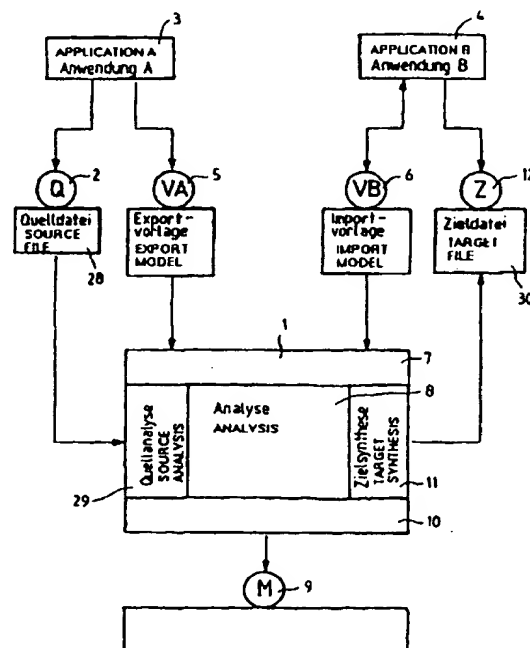
(54) Bezeichnung: VERFAHREN ZUR DATENKONVERTIERUNG

(57) Abstract

The invention concerns a method of converting data between two applications or software packages which use different data formats. A fundamental problem in data-processing today is the combination and exchange of source data from different software packages. Intelligent conversion software predominantly uses data structures of data bases and data-processing equipment. Proceeding from a first recognition of a rough structure, the file form is pre-classified in conversion tools. The conversion software provides an algorithm set by means of which the fine structure is automatically recognized and converted. The analysis of model files of applications A and B is continued as far as the substructure level, the source data finally being converted by means of the analysis into the target data which are imported by application B.

(57) Zusammenfassung

Es wird ein Verfahren zur Konvertierung von Daten zwischen zwei Anwendungen bzw. Softwarepaketen vorgeschlagen, die zueinander unterschiedliche Datenformate benutzen. Als ein Grundproblem der heutigen Datenverarbeitung ist dabei das Zusammenführen und der Austausch von Datenquellen verschiedener Softwarepakete erkannt worden. Für eine intelligente Konvertierungssoftware werden in erster Linie Datenstrukturen von Datenbanken und Datenverarbeitungsanlagen betrachtet. Ausgehend von dem ersten Erkennen einer Grobstruktur wird die Dateiform in Konvertierungstools vorklassifiziert. Die Konvertierungssoftware stellt einen Satz von Algorithmen zur Verfügung, wodurch automatisch die Feinstruktur erkannt und umgewandelt wird. Die Analyse von Vorlagefiles der Anwendungen A und B wird bis in die Substrukturen fortgeführt, wobei letztendlich die Quelldaten mit Hilfe der Analyse in die Zieldaten konvertiert werden, welche von der Anwendung B importiert werden.



LEDIGLICH ZUR INFORMATION

Codes zur Identifizierung von PCT-Vertragsstaaten auf den Kopfbögen der Schriften, die internationale Anmeldungen gemäss dem PCT veröffentlichen.

AL	Albanien	ES	Spanien	LS	Lesotho	SI	Slowenien
AM	Armenien	FI	Finnland	LT	Litauen	SK	Slowakei
AT	Österreich	FR	Frankreich	LU	Luxemburg	SN	Senegal
AU	Australien	GA	Gabun	LV	Letland	SZ	Swasiland
AZ	Aserbaidtschan	GB	Vereinigtes Königreich	MC	Monaco	TD	Tschad
BA	Bosnien-Herzegowina	GE	Georgien	MD	Republik Moldau	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagaskar	TJ	Tadschikistan
BE	Belgien	GN	Guinea	MK	Die ehemalige jugoslawische Republik Mazedonien	TM	Turkmenistan
BF	Burkina Faso	GR	Griechenland	ML	Mali	TR	Türkei
BG	Bulgarien	HU	Ungarn	MN	Mongolei	TT	Trinidad und Tobago
BJ	Benin	IE	Irland	MR	Mauretanien	UA	Ukraine
BR	Brasilien	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Island	MX	Mexiko	US	Vereinigte Staaten von Amerika
CA	Kanada	IT	Italien	NE	Niger	UZ	Usbekistan
CF	Zentralafrikanische Republik	JP	Japan	NL	Niederlande	VN	Vietnam
CG	Kongo	KE	Kenia	NO	Norwegen	YU	Jugoslawien
CH	Schweiz	KG	Kirgisistan	NZ	Neuseeland	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Demokratische Volksrepublik Korea	PL	Polen		
CM	Kamerun	KR	Republik Korea	PT	Portugal		
CN	China	KZ	Kasachstan	RO	Rumänien		
CU	Kuba	LC	St. Lucia	RU	Russische Föderation		
CZ	Tschechische Republik	LI	Liechtenstein	SD	Sudan		
DE	Deutschland	LK	Sri Lanka	SE	Schweden		
DK	Dänemark	LR	Liberia	SG	Singapur		
EE	Estland						

INTERNATIONAL SEARCH REPORT

Inten. nal Application No

PCT/EP 97/01442

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G06F17/30

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 G06F

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	EP 0 449 494 A (IBM) 2 October 1991 see abstract see page 3, column 1, line 1 - page 4, column 4, line 22 ---	1-11
A	US 5 119 465 A (JACK MARTIN L ET AL) 2 June 1992 see abstract; figure 1 ---	1-11
A	EP 0 130 375 A (IBM) 9 January 1985 see abstract see page 2, line 8 - page 2, line 34 -----	1-11

☐ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- * "A" document defining the general state of the art which is not considered to be of particular relevance
- * "E" earlier document but published on or after the international filing date
- * "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- * "O" document referring to an oral disclosure, use, exhibition or other means
- * "P" document published prior to the international filing date but later than the priority date claimed

- * "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- * "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- * "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- * "&" document member of the same patent family

Date of the actual completion of the international search

11 July 1997

Date of mailing of the international search report

22.07.97

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+ 31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+ 31-70) 340-3016

Authorized officer

Katerbau, R

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No.

PCT/EP 97/01442

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
EP 0449494 A	02-10-91	CA 2034905 A JP 4223538 A KR 9616882 B US 5416917 A US 5566332 A	28-09-91 13-08-92 26-12-96 16-05-95 15-10-96
US 5119465 A	02-06-92	NONE	
EP 0130375 A	09-01-85	US 4559614 A JP 1746325 C JP 4034782 B JP 60015739 A	17-12-85 25-03-93 09-06-92 26-01-85

INTERNATIONALER RECHERCHENBERICHT

Int. .onales Aktenzeichen

PCT/EP 97/01442

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES

IPK 6 G06F17/30

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)

IPK 6 G06F

Recherchierte aber nicht zum Mindestprüfstoff gehorende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
A	EP 0 449 494 A (IBM) 2. Oktober 1991 siehe Zusammenfassung siehe Seite 3, Spalte 1, Zeile 1 - Seite 4, Spalte 4, Zeile 22 ---	1-11
A	US 5 119 465 A (JACK MARTIN L ET AL) 2. Juni 1992 siehe Zusammenfassung; Abbildung 1 ---	1-11
A	EP 0 130 375 A (IBM) 9. Januar 1985 siehe Zusammenfassung siehe Seite 2, Zeile 8 - Seite 2, Zeile 34 -----	1-11

☐ Weitere Veröffentlichungen und der Fortsetzung von Feld C zu entnehmen

☒ Siehe Anhang Patentfamilie

* Besondere Kategorien von angegebenen Veröffentlichungen :

* "A" Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

* "E" älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

* "L" Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

* "O" Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

* "P" Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

* "T" Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

* "X" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderscher Tätigkeit beruhend betrachtet werden

* "Y" Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderscher Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

* "&" Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

11. Juli 1997

Absendedatum des internationalen Recherchenberichts

22.07.97

Name und Postanschrift der Internationale Recherchenbehörde
Europäisches Patentamt, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+31-70) 340-3016

Bevollmächtigter Bediensteter

Katerbau, R

INTERNATIONALE RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Internationales Aktenzeichen

/EP 97/01442

Im Recherchenbericht angeführtes Patentdokument	Datum der Veröffentlichung	Mitglied(er) der Patentfamilie	Datum der Veröffentlichung
EP 0449494 A	02-10-91	CA 2034905 A	28-09-91
		JP 4223538 A	13-08-92
		KR 9616882 B	26-12-96
		US 5416917 A	16-05-95
		US 5566332 A	15-10-96

US 5119465 A	02-06-92	KEINE	

EP 0130375 A	09-01-85	US 4559614 A	17-12-85
		JP 1746325 C	25-03-93
		JP 4034782 B	09-06-92
		JP 60015739 A	26-01-85

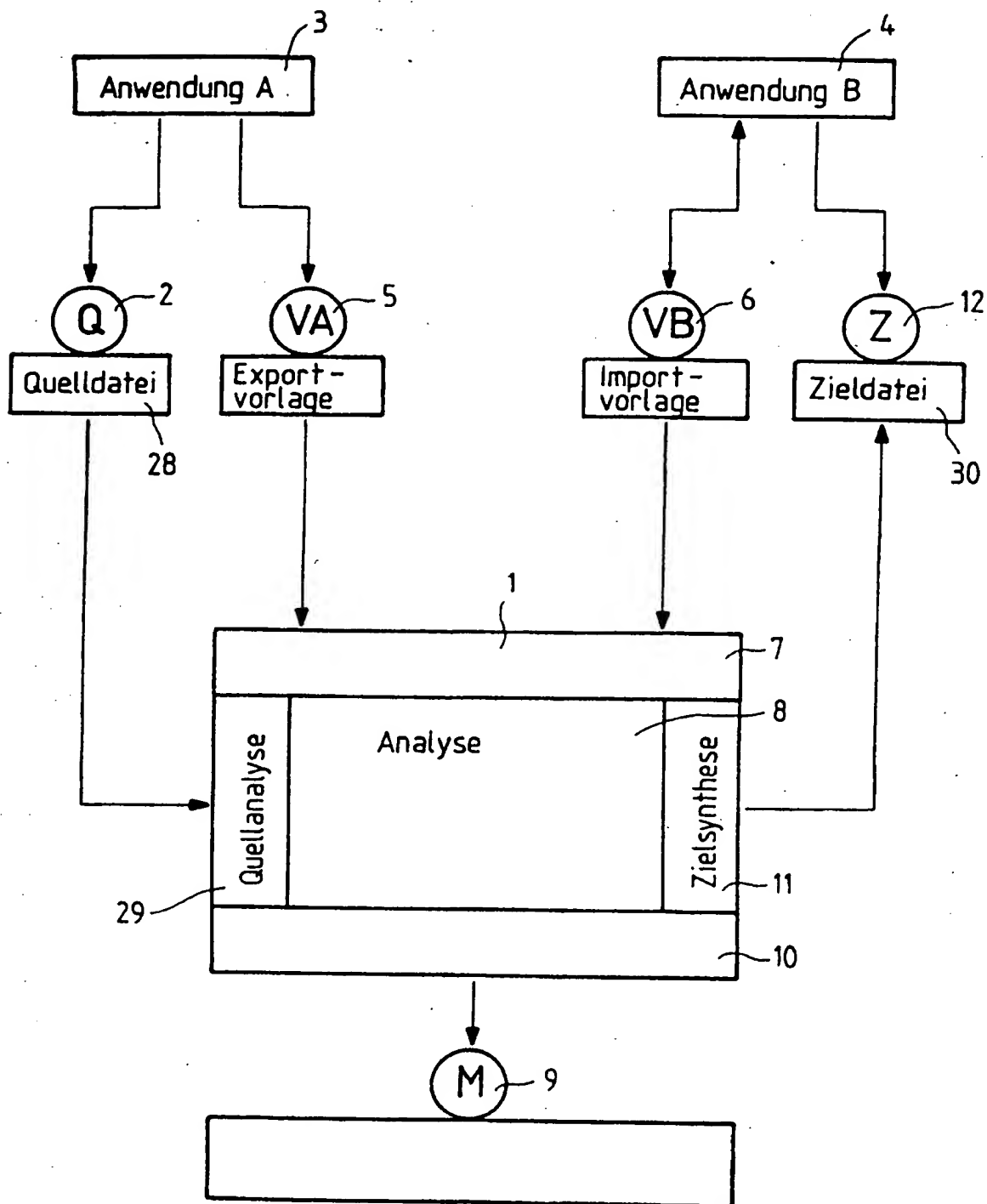


FIG. 1

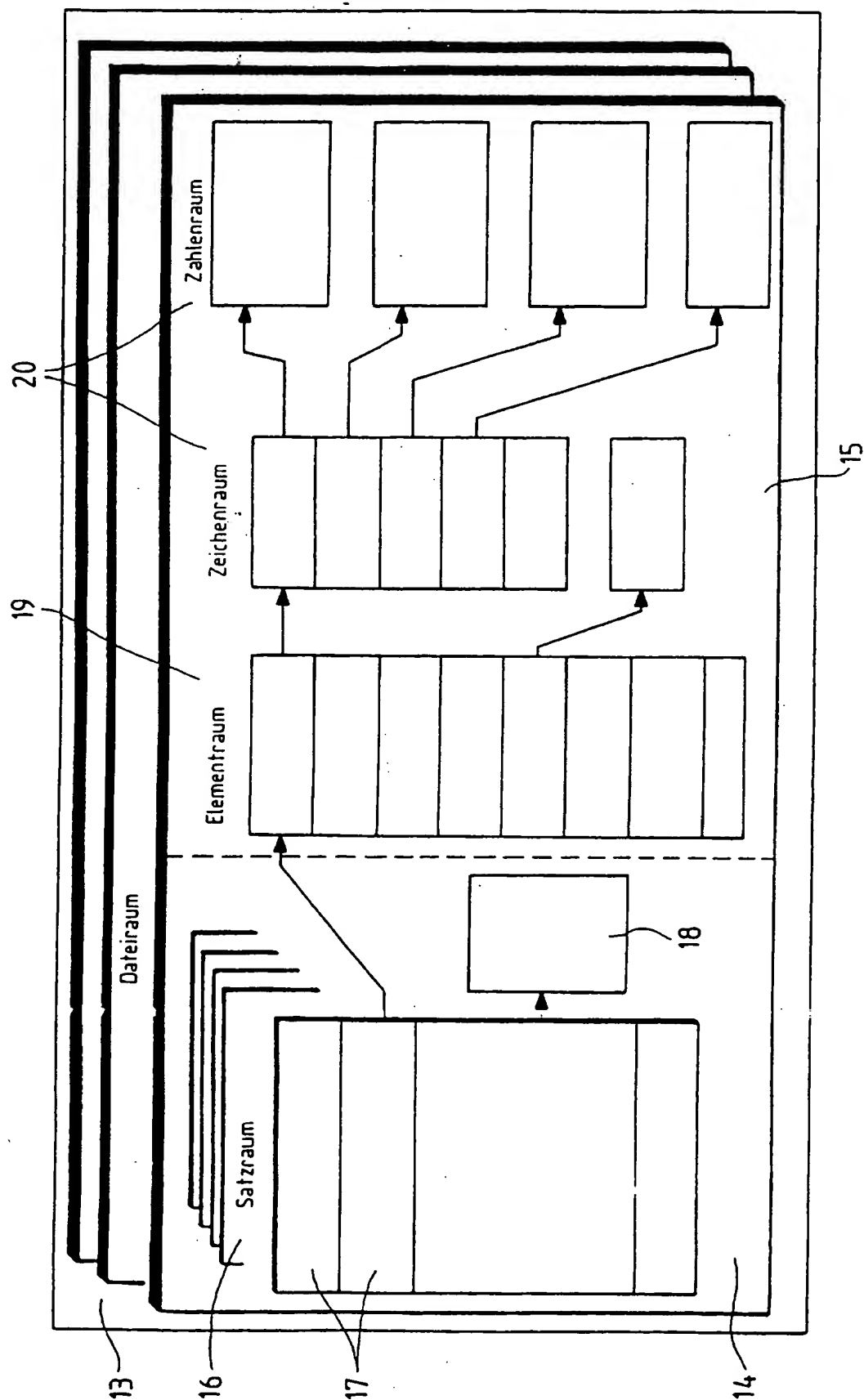


FIG. 2

3/4

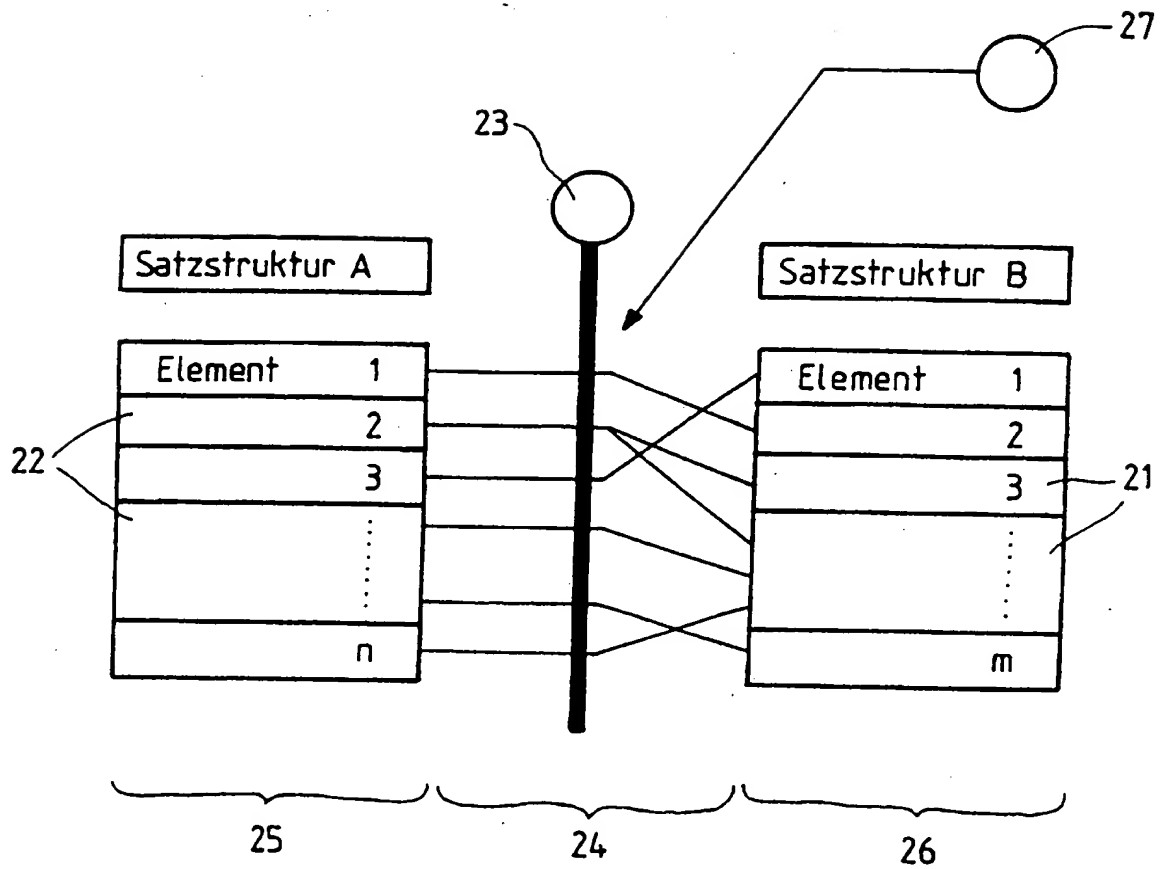


FIG. 3

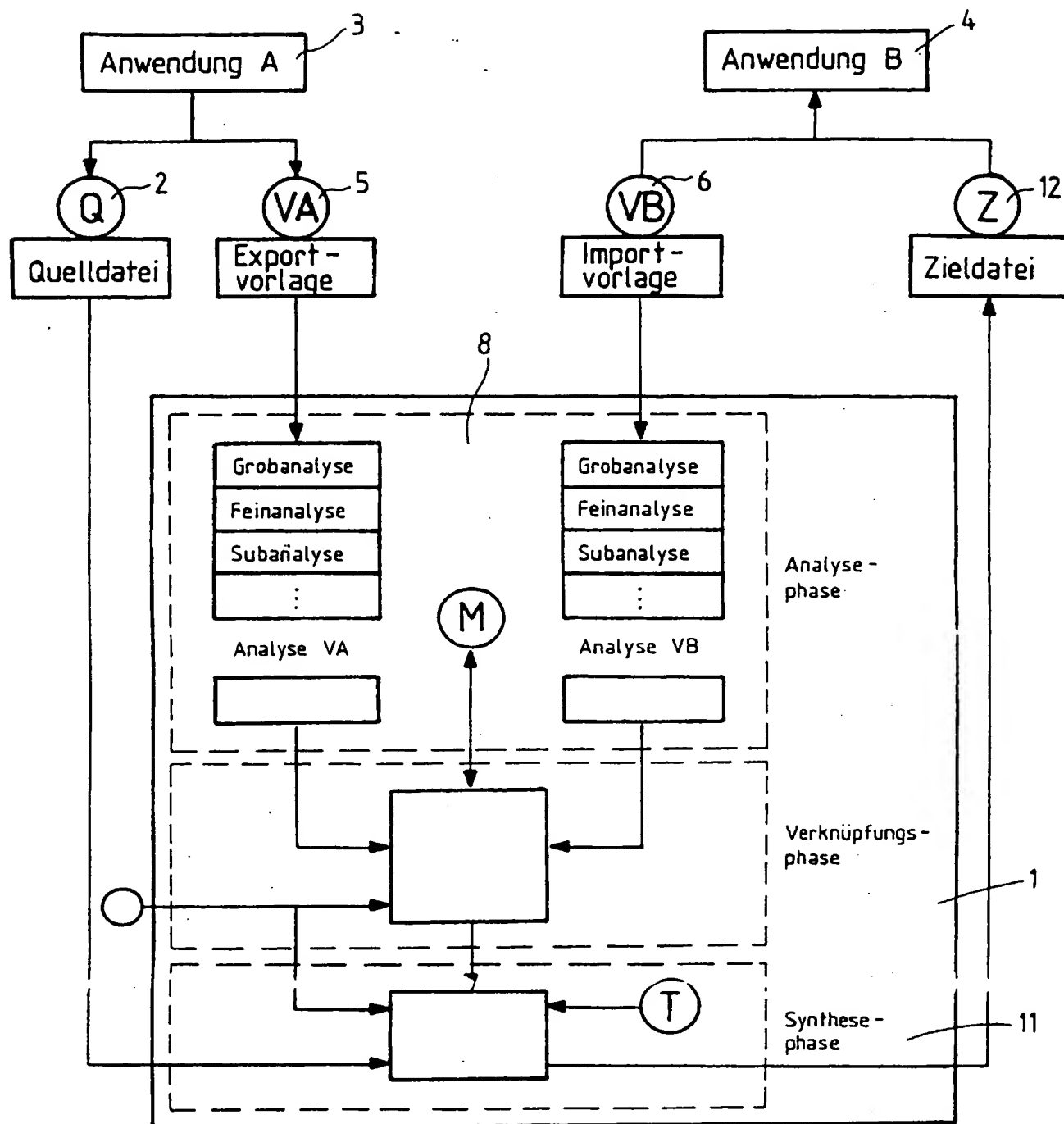


FIG. 4

Beschreibung

5

Verfahren zur Datenkonvertierung

10

Die Erfindung betrifft ein Verfahren zum Konvertieren von Daten zwischen zwei Anwendungen beziehungsweise Softwarepaketen, die unterschiedliche Datenformate benutzen, nach dem Oberbegriff des Patentanspruches 1.

15

Der Erfindung liegt das grundsätzliche Problem der Inkompatibilität von Daten aus verschiedenen Anwendungen bzw. Softwarepaketen zugrunde. Viele Programme besitzen ein eigenes Datenformat für die interne Verarbeitung der Daten. Dieses ist aber nur selten von anderen Programmen zur Weiterverarbeitung einlesbar. Der Anwender steht vor dem Problem, beliebige Daten von Anwendung A nach Anwendung B zu transportieren, wobei ihm im allgemeinen die Datenformate der Anwendungen A und B nicht bekannt sind. Bei den zu transportierenden Daten handelt es sich meist um Datenbanken, es können aber auch im weitesten Sinne Anwendungsdaten auftreten, wie zum Beispiel Textdateien, CAD-Daten, Sound-Formate, Kalkulationsprogramme und andere.

30

Um den vorstehend beschriebenen Datentransfer realisieren zu können, bieten moderne und flexible Programme die Möglichkeit, Daten der gängigsten Formate einzulesen und in ihrem eigenen Format weiterzuverarbeiten. Um einen Austausch zwischen verschiedenen Anwendungen, beispielsweise auch von anderen Betriebssystemen, zu ermöglichen, ist bei einigen Softwarepaketen eine Im- und Exportfunktion im ASCII-Format oder einem anderen gängigen Format vorgesehen.

35

Darüberhinaus gibt es viele Hilfsprogramme, die Datenkonvertierungen von einem Format A nach einem Format B erlauben. Diese haben einen fest vorgegebenen Konvertierungsalgorithmus implementiert, welcher der jeweiligen Anwendung entspricht. Der Anwender kann nun wählen, zwischen welchen Datenformaten A und B das Programm eine Konvertierung durchführen soll. Diese Hilfsprogramme können teilweise die Art der Datei bzw. die Art der Anwendung A automatisch anhand von einfachen spezifischen Merkmalen bei der Dateiform, wie zum Beispiel der Dateiform und festen Infostrukturen und Versionskennungen innerhalb der Datei, erkennen.

Ein spezielles Verfahren zur Datenkonvertierung ist bereits aus der DE 41 28 940 A1 bekannt. Mit Hilfe dieses Verfahrens werden druckaufbereitete digitale Daten eines mit einer Text- und/oder Graphik-Software am Bildschirm einer Datenverarbeitungsanlage erstellten Dokumentes in faksimilespezifische Übertragungsdaten konvertiert. Dieses Verfahren ersetzt die umständliche, zeitaufwendige und fehlerbehaftete Methode, ein am Bildschirm einer Datenverarbeitungsanlage erstelltes Dokument an einem an der Datenverarbeitungsanlage angeschlossenen Drucker auszudrucken und anschließend das auf Papier gedruckte Dokument durch eine optoelektronische Einheit eines Faksimilegerätes abzutasten. Die Abtastdaten der optoelektronischen Einheit werden dann faksimilespezifisch codiert und übertragen.

Sowohl die in den Anwendungen integrierten Konvertierungsprogramme als auch die eigenständigen Hilfsprogramme zur Datenkonvertierung haben beide den Nachteil, daß nur eine begrenzte Auswahl an Datenformaten zur Verfügung gestellt wird, die mit den festen Algorithmen des Programmes konvertiert werden können. Diese Konvertierungen berücksichtigen in der Regel nicht die Elementinhalte und deren Verknüpfungen, wie beispielsweise Zeicheninhalt und Zeichenlänge, sprachliche Aspekte, Existenzregeln, Elementumformungen, Sprachraum, Elementmuster und Elementinterpretation. Außerdem sind die meisten Datenformate

nicht sehr gut standardisiert, und viele Anwendungen bringen ihre eigenen Datenstrukturen ein. Die Datenelemente sind für die Quellanwendung optimiert und können bei der Ziellanwendung Fehler bei der Verarbeitung verursachen, die aufwendig vom Anwender von Hand korrigiert werden müssen oder sogar eine Weiterverarbeitung der Daten unmöglich machen.

Es ist Aufgabe der Erfindung, dem Anwender ein Verfahren zur Datenkonvertierung zur Verfügung zu stellen, das eine flexible Datenangleichung von verschiedenen Anwendungsprogrammen mit relativ einfachen Mitteln ermöglicht.

Diese Aufgabe ist erfindungsgemäß durch die Merkmale des Patentanspruchs 1 gelöst.

Erfinderische Weiterbildungen und Ausgestaltungen der Erfindung sind in den Unteransprüchen beschrieben.

Die erfindungsgemäße Konvertierungssoftware ist dabei in der Lage, Daten eines beliebigen Formates A in Daten eines beliebigen Formates B halb- und auch vollautomatisch zu transferieren, wobei im Prinzip keine Einschränkung bezüglich der möglichen Datenformate auftritt. Da die Datenformate A und B von der erfindungsgemäßen Konvertierungssoftware über die Grobstruktur der Daten hinaus sowohl halb- als auch vollautomatisch in deren Feinstruktur analysiert werden, ist die mögliche Fehlerquote beim Datentransfer von Anwendung A nach Anwendung B deutlich geringer als bei den üblichen Hilfsprogrammen. Zusätzlich sind Datenveränderungen und Datenanpassungen in diversen Ebenen, beispielsweise in der Datei-, Satz-, Elementebene und weitere, vorgesehen, um die Fehlerquote bei der Ziellanwendung zu verringern.

Der Aufbau des erfindungsgemäßen Verfahrens zur Datenkonvertierung ist in der Zeichnung dargestellt und nachstehend erläutert. Darin zeigen:

- Fig. 1 das Informationsablaufdiagramm des Verfahrens zum Datentransfer;
- Fig. 2 ein Diagramm zur Strukturanalyse des Verfahrens mit detailliertem Strukturaufbau;
- 5 Fig. 3 den Syntheseverlauf des Verfahrens für die Bildung der Daten B aus den Daten A;
- Fig. 4 das Informationsablaufdiagramm gemäß Figur 1 in detaillierterer Darstellung.

10 Die Figur 1 zeigt in Form eines Informationsablaufdiagrammes den Ablauf eines Datentransfers mit Hilfe des erfindungsgemäßen Verfahrens bzw. der erfindungsgemäßen Konvertierungssoftware (1). Ein Anwender möchte beliebige Quelldaten Q (2), die von der Anwendung A (3) erzeugt und exportiert werden, zur Anwendung B (4) transportieren, um sie dort weiterzuverarbeiten. Jede der Anwendungen A (3) und B (4) kann ihre Daten in einem für sie selbst verständlichen Format im- und exportieren; dies ist in den Anwendung (3) und (4) implementiert, um Daten zwischen Anwendungen der gleichen Art auf verschiedenen Arbeits-

15 plätzen auszutauschen.

20

Der Anwender kann nun jeweils einen Probeexport von beispielsweise etwa 500 Datensätzen aus dem Datenbestand der Anwendung A (3) und dem der Anwendung B (4) erzeugen. So entstehen das

25 Vorlagefile VA (5) aus Anwendung A (3) und Vorlagefile VB (6) aus Anwendung B (4). Voraussetzung für das Funktionieren des Verfahrens zur Datenkonvertierung ist dabei, daß die Anwendung B (4) ihr eigenes, von ihr erzeugtes, exportiertes Vorlagefile VB (6) wieder importieren kann, was aber selbstverständlich

30 funktionieren sollte. Jetzt können die Vorlagefiles VA (5) und VB (6) mittels der Konvertierungssoftware (1) analysiert werden (7).

Es wäre grundsätzlich auch ein Datenaustausch von Anwendung A

35 (3) zu Anwendung B (4) direkt möglich ohne den Weg über die Datei (zum Beispiel OLE, DDE und weitere) zu gehen. Hier wird

dann also der Umweg über die Dateien gespart, während die Analyse- und Verknüpfungsprozesse natürlich wirksam bleiben sind.

Da der Anwender im allgemeinen nichts über den Aufbau der Datenstrukturen und über die Datenstrukturen der Anwendungen A (3) und B (4) weiß, und darüber auch nichts wissen möchte, sind für die Analyse komplexe Algorithmen erforderlich, um den Anwender möglichst von Entscheidungen zu entlasten und den Analyse- und Verknüpfungsprozeß (8) zu automatisieren oder zumindest zum Teil zu automatisieren. Die implementierten Algorithmen können problemlos erweitert werden, um auch neue, in Zukunft auftretende Datenstrukturen analysieren zu können. Der Umfang der Analyse (8) der Strukturen wird weiter unten anhand von Figur 2 näher erläutert. Nach dem Analyse- und Verknüpfungsprozeß (8) kennt die Konvertierungssoftware (1) die Datenstruktur der Vorlagefiles VA (5) und VB (6) und auch die notwendigen Verknüpfungen zwischen den Elementen der Vorlagefiles VA (5) und VB (6). Dieses Konvertierungsmuster für die Übertragung der Daten von Anwendung A (3) nach Anwendung B (4) wird als Musterablaufdatei M abgespeichert.

Will der Anwender zu einem späteren Zeitpunkt weitere Datentransfers zwischen denselben Anwendungen A (3) und B (4) wie oben erläutert durchführen, so ist es möglich, das eben ermittelte Konvertierungsmuster in Form von Musterablaufdaten M (9) abzuspeichern (10). Somit erübrigt sich für den Anwender beim nächsten Konvertierungsprozeß die Erzeugung der Vorlagefiles VA (5) und VB (6), und die Konvertierungssoftware (1) kann beim Datentransfer ohne Analyseprozeß (8) direkt auf die gesicherten Musterablaufdaten M (9) zurückgreifen und die Musterablaufdaten M (9) wieder einlesen (10).

Wie dazu insbesondere Figur 4 in Ergänzung zu Figur 1 zeigt, unterteilt sich die Analysephase sowohl des Vorlagefiles VA (5) als auch des Vorlagefiles VB (6) in eine Grobanalyse, eine folgende Feinanalyse, eine weitere folgende Subanalyse und so wei-

ter, um schließlich im Ergebnis zur Endanalyse zu gelangen. An die Analysephase schließt sich dann die Verknüpfungsphase an, in welcher die Analysen der Vorlagefiles verknüpft werden.

- 5 In der Synthesephase (11) wird mit Hilfe des ermittelten Analyse- und Verknüpfungsergebnisses eine Konvertierung der Quelldaten Q (2) in die Zieldaten Z (12) erreicht. Im letzten Schritt kann die Anwendung B (4) die Zieldaten Z (12) importieren, da sie in dem von der Anwendung B (4) im- und exportierbaren Datenformat vorliegen, und weiterverarbeiten.

Bei diesem erfindungsgemäßen Verfahren zur Datenkonvertierung ist es für den Benutzer zudem möglich, durch Verwendung von vom Benutzer definierten Sonderregeln bei der Synthese (11) der
15 Quelldaten Q (2) in die Zieldaten Z (12) Datenveränderungen vorzunehmen. Als Beispiele, und keineswegs vollständig aufgezählt, seien hier nur der Ersatz von ä, ö, ü durch ae, oe ue und die Umbildung der Elemente [Sehr geehrter] und [Herr xyz] in die Elemente [Sehr geehrter Herr] und [xyz] genannt. Es sind
20 hierbei natürlich auch sehr komplexe Datenveränderungen möglich, die am besten mit dem Begriff des "Morphing" aus der Bildverarbeitung verglichen werden können, der die Wandlung von einem Muster A nach Muster B ohne Informationsverluste beschreibt.

- 25 Wie Figur 1 zeigt, wird die zu exportierende Quelldatei (28) der Anwendung A (3) in die Quellanalyse (29) gegeben. Nach erfolgter Analyse für automatische Regelbildung und Verknüpfungen im Block (8) wird die Zielanalyse (11) erstellt. Die konvertierten Daten werden sodann in die importierbare Zieldatei (30) gegeben. Bei dem Vorgang des Exportierens der Quelldatei (28) in die Zieldatei (11) mittels der Musterablaufdaten M ist eine Quell- oder Zielanalyse nicht mehr erforderlich. Auch die Regelbildung und die Verknüpfungen wurden in diesem Fall bereits
35 mit den Vorlagefiles VA (5) und VB (6) erzeugt. Bei der Synthese wird die Quelldatei Q (2) gelesen und mit dem Datenfilter

aus den Musterablaufdaten M bearbeitet. Die generierten Datensätze werden in die Zieldatei geschrieben. Das Datenfilter kann sich externer Dateien zum Abgleich wie auch interner Dateien zum Suchen und Sortieren bedienen.

5

In Figur 2 ist ein Diagramm dargestellt, welches den Struktur-
aufbau einer allgemeinen Datentransferdatei in detaillierter
Form zeigt. Die Analyse (8) der Vorlagefiles VA (5) und VB (6)
unterteilt sich in mehrere Phasen. Ausgehend von dem ersten Er-
10 kennen der Grobstruktur (13) wird die Dateiform wie bei bisher
bekannten Konvertierungsprogrammen zunächst vorklassifiziert.
Die intelligente Konvertierungssoftware stellt nun zusätzlich
einen Satz von Algorithmen zur Verfügung, der in der Lage ist,
15 automatisch die Feinstruktur (14) der vorliegenden Daten zu er-
kennen. Die Analyse (8) wird fortgeführt, um auch die Substruk-
turen (15) innerhalb der Datenelemente (17) zu erkennen.

Wie auch gut erkennbar aus Figur 3 hervorgeht, dient die Grob-
struktur (13) dem Erkennen der Filestruktur, der Trennzeichen,
20 der Elementbezeichner, der Zeichenlänge usw. im Dateiraum und
im Satzraum (16). In der Grobstruktur (13) werden Dateien nach
ihrer Dateiform unterschieden. Die Unterschiede sind zum Bei-
spiel anhand von Dateiendungen oder Infostrukturen innerhalb
der gesamten Datei zu erkennen.

25

Die Feinstruktur (14) dient dem Erkennen von Elementexistenz,
Element-Zeicheninhalten, Elementzeichenlänge usw. wie Elemen-
traum (19) und Zeichenraum (20). Die Feinstruktur (14) bezieht
sich auf die Satzebene, die den Aufbau eines einzelnen Daten-
30 satzes im sogenannten Satzraum (16) aus den Elementen (17),
sowie die Verknüpfungen der Satzelemente (17) untereinander,
den sogenannten Satzregelraum (18) beinhaltet. Bei der Merk-
malsfindung für den Datensatzaufbau und die Datensatzstruktur
werden beispielsweise Feldnamen in einem Datensatz (16) analy-
35 siert, die Verwendung spezieller Trennsymbole für Elemente (17)
und Datensätze (16), die Verwendung von Start- und Stopsymbolen

für die Strukturierung innerhalb der Elemente (17) und die Anzahl der Elemente (17) pro Datensatz (16) ermittelt.

Die Substruktur (15) dient dem Erkennen von Worten und Subelementen mit Sub-Trennern in den einzelnen Elementen, wie Elementraum (19), Zeichenraum, Wortraum. Die Substruktur untergliedert sich also nochmals in die Elementebene, dem Elementraum (19), und die Zeichenebene, dem Zeichenraum mit Zahlenraum (20). In der Elementebene (19) werden die Existenzmerkmale für ein Element (17) festgelegt.

Als Existenzmerkmale treten zum Beispiel auf: minimale und maximale Datenelement-Zeichenlänge, Datenelement-Existenzvorgaben (muß, kann, darf), zulässige und vorkommende Zeichen in einem Datenelement (17), verwendeter Sprachraum, begrenzte Listenstrukturen innerhalb von Datenelementen (17), Wortanalysen, etc.. Die Zeichenebene (20) gibt die möglichen Auswahlräume für Zahlen, Buchstaben, Sonderzeichen, OEM-Zeichen und andere vor.

So kann zum Beispiel das Element "Arb.Zeit" einen

1. Satz: "Meier/Müller/Graf/Kemper", einen
2. Satz: "8h /7H /9h /8h" und einen
3. Satz: "5H /9h /11h /7h" enthalten.

Demzufolge gibt es in dem Element "Arb.Zeit" offenbar Substrukturen, die durch das Zeichen "/" getrennt sind. Solche und ähnliche Substrukturen werden erkannt und bei der Analyse wie zusätzliche Elemente mit Bezeichner, nämlich Meier, Müller, Graf, Kemper, und entsprechenden Inhalten behandelt.

30

Wie oben an anderer Stelle bereits ausgeführt, enthält das Element "Anrede" mehrere Wörter:

1. Satz: "Sehr geehrter Herr",
2. Satz: "Sehr geehrte Frau",
- 35- 3. Satz: "Hallo, Herr" und
4. Satz: "Sehr geehrter Herr Dipl.-Ing".

Diese Substrukturen werden als Wörter in Wortlisten übernommen und gegebenenfalls einzeln weiter nach zusätzlichen Mustern und Regeln analysiert.

5 Zur Verdeutlichung dient ein weiteres Beispiel. Das Element "Strasse" enthält die Muster

1. Satz: "Hofweg 13",
2. Satz: "Am langen Kant 77" und
3. Satz: "Bremer Strasse 19".

10

Diese Substrukturen (15) werden als Worte analysiert und nach Mustern untersucht. Es können so zum Beispiel Strassenname und Hausnummer gesplittet werden.

15 Von großer Wichtigkeit ist es, daß sich die Substrukturen (15) auf Objekte in einer Ebene unter den Satzelementen beziehen. Hier werden die Elementinhalte nach weiteren Mustern abgesucht.

20 Bei der Musterstruktur werden mit den vorgenannten Analyse- Ergebnissen die Muster- und Verknüpfungsanalysen durchgeführt (Satzraum (16), Satzregelraum, Elementraum (19), Wörteraum). Dabei werden übergreifende Muster und Regeln mittels fester Algorithmen und neuronaler sowie Fuzzy-Algorithmen bestimmt.

25 Um die Mustererkennung bei der sprachlichen Analyse (8) von Feinstruktur (14) und Substruktur (15) möglichst automatisch ablaufen lassen zu können, ist bei den Algorithmen die Verwendung moderner Techniken, wie neuronale Netze und Fuzzy-Logik neben festen Algorithmen, erforderlich. Läßt man solche Ansätze
30 weg, so muß der Anwender eventuell einzelne Verknüpfungen von Hand durchführen.

Die Figur 3 zeigt in Diagrammform nun den Syntheseverlauf (11) für die Bildung der Datenelemente (21) für die Anwendung B (4)
35 aus den Datenelementen (22) der Anwendung A (3). Aus der oben beschriebenen Analyse (8) der Vorlagefiles VA (5) und VB (6)

ergibt sich ein Datenfilter (23), der die zuvor ermittelten Verknüpfungsregeln (24) zwischen den Datenelementen (21) und (22) enthält. Schickt man die Datensätze (25) aus der Anwendung A (3) durch diesen Datenfilter (23), so erhält man neue Datensätze (26) mit neuen Datenelementen (21), die in einem von der Anwendung B (4) im- und exportierbaren Datenformat vorliegen. Die Anzahl der Datenelemente (22) der Datensätze (25) aus der Anwendung A (3) ist nicht unbedingt gleich der Anzahl der Datenelemente (21) der Datensätze (26) für die Anwendung B (4), die Anzahl der Datensätze (25) und (26) ist im allgemeinen für beide Anwendungen A (3) und B (4) gleich.

Zusätzlich zu dem aus der Analyse (8) der Vorlagefiles VA (5) und VB (6) entstandenen Datenfilter (23) für den Datentransfer von Anwendung A (3) nach Anwendung B (4) ist auch ein Abgleich der Datenelemente (21) und (22) mit einer externen Datenbank oder mehreren externen Datenbanken (27) möglich, beispielsweise auch die Mischung von mehreren Datenbanken, wie Kundendatei + Rechnungsdatei. Mit einem solchen externen Abgleich können zum Beispiel die Postleitzahlen der Datensätze (25) und (26) kontrolliert werden, indem man die Datenelemente (21) und (22) von Land, Stadt, Straße und Postleitzahl mit denen der externen Datenbank (27) vergleicht.

Zusätzlich ist das Anlegen einer oder mehrerer interner Zwischendatenbanken zum Suchen und Sortieren auf Dateiebene möglich. Dies ist erforderlich, um ein schnelles Suchen und Sortieren, beispielsweise nach dem Alphabet oder speziellen Merkmalen, zu ermöglichen.

30

35

Bezugszeichenliste

5	1	Konvertierungssoftware bzw. Verfahren
	2	Quelldaten Q
	3	Anwendung A
	4	Anwendung B
10	5	Vorlagefile VA
	6	Vorlagefile VB
	7	Eingang Analysevorlagen
	8	Analyse, Analyseprozeß
	9	Musterablaufdaten M
15	10	Speichern der Musterablaufdaten
	11	Synthese, Synthesephase
	12	Zieldaten Z
	13	Grobstruktur
	14	Feinstruktur
20	15	Substruktur
	16	Datensatz
	17	(Daten-) Element
	18	Satzregelraum
	19	Elementebene
25	20	Zeichenebene
	21	Datenelemente
	22	Datenelemente
	23	Datenfilter
	24	Verknüpfungen, Verknüpfungsregeln
30	25	Datensätze
	26	Datensätze
	27	externe Datenbank
	28	Quelldatei
	29	Quellanalyse
35	30	Zieldatei

5

Patentansprüche

10 1. Verfahren zur Konvertierung von Daten und nicht-datei-bezo-
genen Methoden zwischen zwei Anwendungen bzw. Softwarepaketen,
die zueinander unterschiedliche Datenformate benutzen,
dadurch gekennzeichnet,
daß aus dem Datenbestand der ersten Anwendung A (3) und dem Da-
15 tenbestand der zweiten Anwendung B (4) jeweils ein Probe-Export
von Datensätzen relativ geringer Menge erzeugt wird, aus denen
ein Vorlagefile VA (5) aus Anwendung A (3) und ein Vorlagefile
VB (6) aus Anwendung B (4) gebildet werden, von denen die An-
wendung B (4) das von ihr erzeugte Vorlagefile VB (6) wieder
20 importieren kann, daß die Vorlagesfiles VA (5) und VB (6) zur
Ermittlung ihrer Datenstrukturen und gegebenenfalls zur Ermitt-
lung der Verknüpfungen zwischen den Elementen durch eine Kon-
vertierungssoftware (1) analysiert werden, und daß die Quell-
daten Q (2) mit Hilfe der Analyseergebnisse in die Zieldaten Z
25 (12) konvertiert werden, welche von der Anwendung B (4) impor-
tiert werden.

2. Verfahren nach Anspruch 1,
dadurch gekennzeichnet,

30 daß das Analyse-Ergebnis als Konvertierungsmuster für die Über-
tragung der Daten des ersten Anwenders A (3) zum zweiten Anwen-
der B (4) in Form von Musterablaufdaten M (9) abgespeichert
wird.

35 3. Verfahren nach Anspruch 1,
dadurch gekennzeichnet,

daß die Quelldaten Q (2) nach Durchlauf der Analyse (8) und Verknüpfung in eine Zielsynthese (11) gegeben werden, von der sie in die Zieldaten Z (12) konvertiert werden.

- 5 4. Verfahren nach einem der Ansprüche 1 bis 3,
dadurch gekennzeichnet,
daß die Analyse (8) der Vorlagefiles VA (5) und VB (6) in
mehrere, insbesondere drei Phasen unterteilt ist, wobei nach
dem Erkennen der Grobstruktur (13) die Dateiform zunächst
10 vorklassifiziert wird, anschließend durch die Konvertierungs-
software ein Satz von Algorithmen zur automatischen Erkennung
der Feinstruktur (14) zu Verfügung gestellt wird und schließ-
lich in Fortführung der Analyse die Substrukturen (15) inner-
halb von Datenelementen (17) erkannt werden.
- 15 5. Verfahren nach Anspruch 4,
dadurch gekennzeichnet,
daß in der Grobstruktur (13) die Dateien nach ihrer Dateiform
unterschieden werden.
- 20 6. Verfahren nach Anspruch 4,
dadurch gekennzeichnet,
daß sich die Feinstruktur (14) auf die Satzebene bezieht, die
den Aufbau eines Satzes (16) aus Elementen (17) und die Ver-
25 knüpfungen der Satzelemente (17) untereinander beinhaltet.
7. Verfahren nach Anspruch 4,
dadurch gekennzeichnet,
daß die Substruktur (15) sich in die Elementebene (19) und in
30 die Zeichenebene (20) mit Zeichenraum und Zahlenraum unter-
gliedert, wobei in der Elementebene (19) die Existenzmerkmale
für ein Element (17) festgelegt werden.
8. Verfahren nach einem der vorgenannten Ansprüche,
35- dadurch gekennzeichnet,
daß bei den Algorithmen für die Mustererkennung der sprachli-

chen und logischen Analyse (8) die Technologien der neuronalen Netze und/oder der Fuzzy-Logik eingesetzt werden.

9. Verfahren nach einem der vorgenannten Ansprüche,
5 dadurch gekennzeichnet,
daß aus der Analyse (8) der Vorlagefiles VA (5) und VB (6) ein
Datenfilter (23) gebildet wird, der die Verknüfungsregeln (24)
zwischen den Datenelementen (21,22) auf der Dateiebene oder zu-
sammen mit einer oder mehreren externen Datenquellen (Abgleich)
10 aufweist, wobei nach Durchlauf von Datensätzen (25) aus der An-
wendung A (3) durch den Datenfilter (23) neue Datensätze (26)
mit neuen Datenelementen (21) erhalten werden, die in einem von
der Anwendung B (4) im- und exportierbaren Datenformat vor-
liegen.

15
10. Verfahren nach einem der vorgenannten Ansprüche,
dadurch gekennzeichnet,
daß die Datenelemente (21,22) mit einer externen Datenbank (27)
abgleichbar sind.

20
11. Verfahren nach einem der vorhergehenden Ansprüche,
dadurch gekennzeichnet,
daß eine oder mehrere interne Zwischendatenbanken zur Suche und
Sortierung von Datensätzen eingerichtet wird bzw. werden.

25

30

35